

Personalized Recommendation Model: An Online Comment Sentiment Based Analysis

S. Chen, X. Lv, J. Gou

Songmin Chen

School of Economics and Management
Beijing Jiaotong University
Beijing 100044, China
18120608@bjtu.edu.cn

Xiyan Lv

School of Economics and Management
Beijing Jiaotong University
Beijing 100044, China
lvxiyan@bjtu.edu.cn

Juanqiong Gou*

School of Economics and Management
Beijing Jiaotong University
Beijing 100044, China
*Corresponding author: jqgou@bjtu.edu.cn

Abstract

Traditional recommendation algorithms measure users' online ratings of goods and services but ignore the information contained in written reviews, resulting in lowered personalized recommendation accuracy. Users' reviews express opinions and reflect implicit preferences and emotions towards the features of products or services. This paper proposes a model for the fine-grained analysis of emotions expressed in users' online written reviews, using film reviews on the Chinese social networking site Douban.com as an example. The model extracts feature-sentiment word pairs in user reviews according to four syntactic dependencies, examines film features, and scores the sentiment values of film features according to user preferences. User group personalized recommendations are realized through user clustering and user similarity calculation. Experiments show that the extraction of user feature-sentiment word pairs based on four syntactic dependencies can better identify the implicit preferences of users, apply them to recommendations and thereby increase recommendation accuracy.

Keywords: online review, sentiment analysis, feature-sentiment word pairs, personalized recommendation.

1 Introduction

With the popularity of the Internet and the integration of the web into people's daily lives, the number of people online continues to grow. According to the statistical report on the development of China's Internet, by June 2019, China had 854 million Internet users, 61.2% Internet penetration rate, and 847 million mobile Internet users, up 29.84 million from the end of 2018, 99.1% of Internet users use mobile phones to surf the Internet [22].

Regarding the scale of online video use, the number of users in China reached 759 million or 88.8% of all Chinese Internet users. Each major video platform further subdivides into content categories with specialized production and operation based on core product types such as TV series, movies, variety shows and animation. The platforms are continuing to expand into emerging product types such as games, esports and music.

When users go online, they leave behavioral information through browsers and other applications. Through the mining of online user data such as purchase, browsing and rating behaviour, user preferences can be analyzed and applied to the personalized recommendations of goods or services. This method, however, does not consider the user's online reviews.

With the development of web3.0, users can express their real opinions and preferences and are keen to do so, as seen in the increasing amount of evaluation information about products or services available on the Internet. This evaluation information will often affect other users' selection decisions. Users' reviews usually contain their subjective opinions and reflect their preferences and emotional tendencies towards the attributes of products or services. By analyzing users' online reviews, we can discover not only the features of products and services and their advantages and disadvantages, but also users' preferences and concerns.

By combining the information in users' online reviews with their basic personal information and online behavioral characteristics, the problems of sparse user data and system cold starts in the recommendation process can be better addressed and more accurate personalized recommendations can be realized.

2 Literature review

2.1 Online review

User reviews reflect users' sentiments and behavioral decisions [6] and contain users' relevant knowledge, which will add to "Internet word-of-mouth" and have an important impact on other users' purchasing behaviors. At present, the research on the influence of online reviews on user behavior mainly focuses on the two aspects of "the total number of reviews" and "the polarity of reviews". The total number of reviews reflects the "heat" of goods or services, that is, the degree of hot sales. The more the total number of reviews, the more likely it is to arouse users' potential interest and attention. Some studies have shown that the number of online reviews on films is positively correlated with the final box office sales [2]. The "polarity" of reviews refers to the "positive" or "negative" nature of reviews, also known as "sentiment polarity" and "directional nature of reviews", which reflects customers' satisfaction with goods or services [8].

In 2007, A. Ghose and P. G. Ipeirotis [9] firstly analyzed the influence of subjective and objective tendency and the degree of mixing of subjective and objective tendency on the usability of online reviews for search-type products. In 2010, S. M. Mudambi [15] studied the influence of review polarity, review depth and commodity type on perceived review usefulness by establishing a user evaluation usefulness model. Through numerical simulation, Chen J et al. [4] studied how potential customers form their own opinions under the combined effect of positive and negative reviews. The results showed that users at middle and low levels are often persuaded by online reviews, regardless of their initial opinions on products. However, online reviews have less impact on consumers with higher membership levels, who often make purchase decisions based on their first impressions of the product.

2.2 Sentiment analysis

In the related researches at home and abroad, emotion analysis methods can be generally classified into three categories: emotion dictionary-based method, machine learn-based method and linguistic method. The method based on emotion dictionary to judge the polarity of the pending words, which can be regarded as the polarity matching task [17]. Emotion analysis in recommendation systems mainly includes the extraction of attribute words and emotion words and the prediction of attribute words' score [12, 13, 18]. Some literature [12] describes the mining of attribute words and emotion words in users' online reviews based on association rules, taking nouns or noun phrases with high frequency as attribute words. Other literature [20] mentions the extraction of attribute words and emotion words from the review text, and then predicts users' ratings of item attributes based on statistical methods. The approach based on linguistics is mainly to extract the syntactic features of words or phrases through linguistics. Additional literature [1] uses word dependence to show how the emotions of a single phrase affect the emotions of a whole sentence.

The key in machine-learning-based sentiment analysis is to construct a sentiment polarity classifier through a training set and then the classification of subjective reviews [19]. This method does not rely entirely on dictionary or corpus, but relies on the training effect of the classifier, and needs more time to complete the annotation of training data than the dictionary or corpus-based method. Manek et al. [14] proposed a feature selection method for SVM classifier based on the Gini coefficient. Each word in the text is not equally important to the sentiment analysis of the text, and the computer cannot automatically judge the importance of the word. Deng et al. [5] proposed a supervised weight assignment scheme based on the importance of the word in the whole text and the importance of expressing emotion. Review texts lack logicity and most texts are disordered. The supervision of the general learning algorithm accuracy is lower when dealing with disordered text. Perikos [16] designed an integrated classifiers based on three classifiers: the first and the second are statistical (Naive Bayes and maximum entropy) and the third one is a knowledge-based tool performing deep analysis of the natural language sentences.

2.3 Recommendation system

Current recommendation systems can be divided into two categories: content based recommendation and collaborative filtering recommendation. A content-based recommendation system [3] looks for similar items based on the similarity of content browsed by different users. Its advantage is simple and effective, but its disadvantage is that it cannot mine the information that new users are interested in.

Collaborative filtering is the most common recommendation system. It was first proposed by Goldberg and other scholars in 1992 and implemented in the Tapestry experimental email system, which enabled users to extract email lists [10]. At present, collaborative filtering is mainly divided into either user-based or project-based collaborative filtering. The basic idea is to use scores to calculate user or product similarity, find the nearest neighbor collection and then make recommendations. Collaborative filtering is a hot research topic in the academic circle. Traditional collaborative filtering methods have disadvantages such as cold starts, sparse data, large amount of data and poor scalability. In order to better solve these problems, scholars are now focusing on the improvement of traditional collaborative filtering.

Wei Wang and Hongwei Wang [21] have applied sentiment analysis technology to recommendation systems, calculated similarity through analyzing the preference polarity (positive and negative) of users for product features and improved the accuracy of the recommendation algorithm by increasing the precision of the nearest set of users. Gan [7] divided the user's sentiment into positive, negative, neutral and conflict types when analyzing the user's review text. They then optimized the collaborative filtering score prediction model according to the emotional tendency expressed by the users to improve the accuracy of the recommendation algorithm.

3 Recommendation model construction based on sentiment analysis

The recommended process in this article can be described in the following steps, with specific research ideas shown in figure 1.

- Step 1. Get the user ID, user review, item ID, and item score from the user's online reviews. Define N users $U = (u_1, u_2, u_3, u_4 \dots u_n)$, Set $D = (d_1, d_2, d_3 \dots d_m)$ that represent M review texts for K items $T = (t_1, t_2, t_3 \dots t_k)$.
- Step 2. By using syntactic relation rules, extract sentiment words from each online review and use modified characteristic words to form the feature-sentiment word pairs.
- Step 3. According to the feature words extracted from the feature-sentiment word pairs, find the potential features of L items $F = (f_1, f_2 \dots f_l)$, and then calculate the user's attention λ to each feature of the item.
- Step 4. Calculate the sentimental score of feature words in users' online reviews, calculate the similarity between users, find out the nearest neighbor and make recommendations.

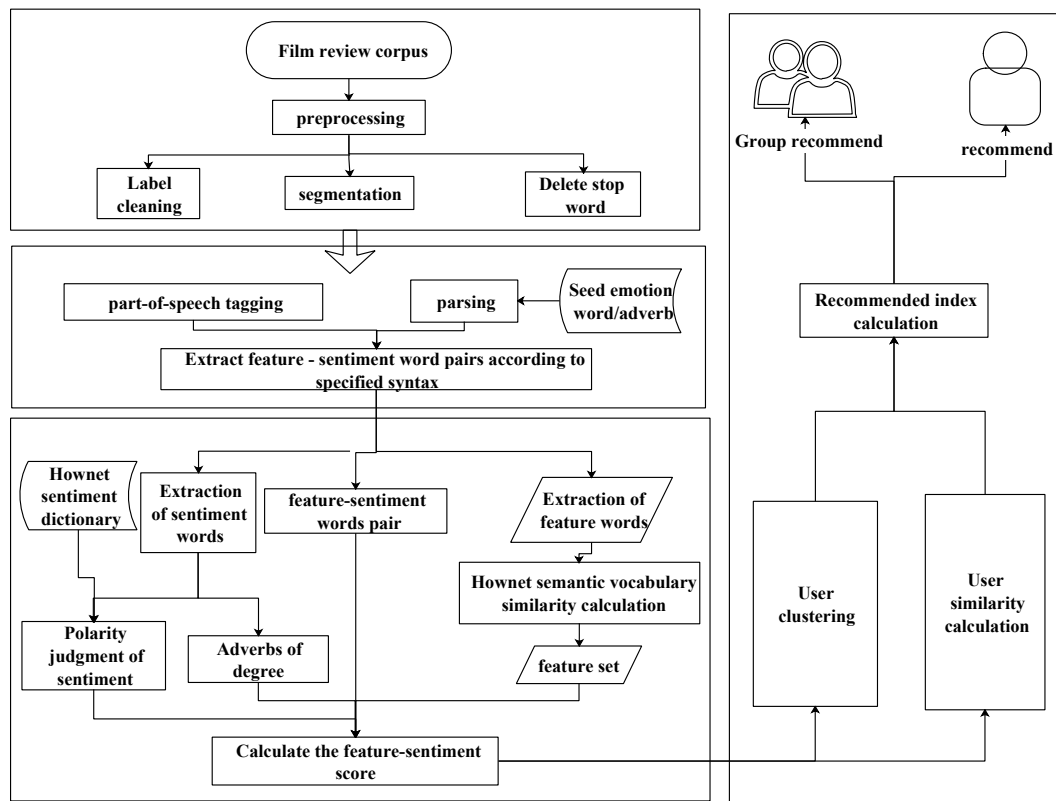


Figure 1: Schematic diagram of research ideas

3.1 Commodity information acquisition

In order to verify the actual effect of sentiment analysis on personalized recommendations, this paper adopts the Scrapy method based on the Python programming language to analyze data on the Chinese movie social networking site Douban, which mainly includes user ID, Douban ID, user review and user score. The collected data set includes a total of 215,032 reviews, deletes users with less than 10 movies and all movies with less than 10 reviews in the data table, for a final total of 67376 reviews on 1464 movies by 2283 users.

3.2 Film review processing based on feature-sentiment word pair extraction

In order to explore the features of movies described by users and words expressed towards evaluation objects, this paper proposes a syntactic dependency-based extraction method for sentiment words. This method is mainly composed of two parts: (1) natural language processing technology is used to process the film review corpus, such as word segmentation, part-of-speech tagging, and syntactic analysis. The syntactic dependence between sentiment words and evaluation objects is explored. (2) Use the syntactic dependency rules constructed in the first step to do pattern matching to identify the evaluation objects and sentiment words in the review sentences, to form feature-sentiment word pairs, and extract the feature-sentiment word pairs whose occurrence frequency exceeds 12. Words describing film features are the evaluated objects in the review data. Sentiment words, or opinion words, are the words used in the review text to express users' positive or negative feelings towards the features of the item. For example, in the review "the picture is good", "the picture" represents the feature of the film, and the word "good" expresses a positive sentimental attitude.

Hu used the two syntactic dependencies of subject-verb relation and attribute relation to extract feature words and sentiment words [11] without considering the verb object relationship and the adverbial relationship, resulting in reduced accuracy in the result. Therefore, in this paper, the seed word bank of sentimental words was added in the process of extracting feature words, and adverbial relation and adverb-object relation were added in the whole process of extraction. Finally, this paper considers four kinds of dependency relations in the process of feature-emotion word pair extraction: (1) subject-verb relation (SUV); (2) verb-object relation (VOB); (3) attribute relation (ATT); (4) adverbial relation (ADV). The review data is filtered according to these four dependencies. This method does not need the seed domain feature set, directly extracts real and reliable evaluation object and sentiment word information and does not require any domain knowledge base except the review data set of the item. In addition, it can be applied with good portability to any domain sentiment analysis if the new domain review data is provided. The extraction results of feature words and sentiment words are shown in table 1.

Table 1: Preprocessing table

relations	Processing content	Output
SUV	The sentence after the participle	image quality/ not high/
SUV	Extract feature-sentiment word pairs	<image quality, not high >
VOB	The sentence after the participle	He/send/a flower/to/her
VOB	Extract feature-sentiment word pairs	<send, flower>
ATT	The sentence after the participle	The movie/ has/ a perfect/ plot/
ATT	Extract feature-sentiment word pairs	<plot, perfect>
ADV	The sentence after the participle	Like /the movie /very much
ADV	Extract feature-sentiment word pairs	< movie, like very much >

3.3 Movie feature clustering based on dictionary

After the extraction of online review feature-sentiment word pairs, the corresponding noun set was established by using Chinese frequent noun non-item feature rules. The extraction results were filtered from the perspective of Chinese semantic and grammatical knowledge and the frequent words irrelevant to film evaluation were removed. In this paper, a HowNet-based semantic vocabulary similarity calculation method is used to cluster the feature words in the table. The HowNet algorithm is shown in formula 1, where $Dis(w_1, w_2)$ represents the distance between the feature words w_1 and w_2 , while $Sim(w_1, w_2)$ is the similarity between the feature words w_1 and w_2 , and the parameter γ can be adjusted according to the actual demand. When $Sim(w_1, w_2)$ is higher than the threshold value, w_1 and w_2 are grouped under the same feature.

$$sim(w_1, w_2) = \frac{\gamma}{Dis(w_1, w_2) + \gamma} \quad (1)$$

With the help of the expanded HowNet dictionary, the similarity among feature words was calculated. According to the similarity results, HowNet classified the candidate feature words into different

Table 2: Attribute features - feature word distribution table

Attribute feature	Feature word
F1 plot	background, story, play, theme
F2 emotion	emotion, touching, healing
F3 figure	figure, woman, man, actor, actress, movie king, chief actor, leading lady, character
F4 acting skill	acting skill, acting, expression, actor's lines, language
F5 staff behind the scenes	director, scriptwriter, author
F6 techniques of expression	narrative, narration, interposed
F7 plot development	transition, paragraph bedding, paragraph
F8 ending	ending, finish, outset, surprise
F9 impressions	Film quality, visual effect, photography, a full-length shot, scenery, film editing, scene, exterior, shoot
F10 sound effect	ending, finish, outset, surprise
F11 embedded value	inspirational, encourage, edification, thinking, righteousness
F12 early film	old film, time, classics, history, reminiscence
F13 affectional film	love, affectional film, first love, romance
F14 literary film	literature and art, youth, warm and sweet
F15 humanity	life, society, humanity, live
F16 youth film	Maiden, young boy, growth, dream
F17 Dracula movie	Dracula movie, terror, ghost, suspense, thriller
F18 political film	politics, social class, hierarchy
F19 war file	war, revolution, disaster, violence, bloodiness
F20 science fiction film	science fiction, science fiction film, magic, fairyland, alien, robot
F21 film adapted from a play	recompose, fiction, original, masterwork
F22 comedy	comedy, funnyman, humor, Chaplin
F23 Chinese film	China, Chinese Mainland, Taiwan, Hongkong and Taiwan
F24 American film	Europe and America, Europe, America, foreign country, American, Hollywood
F25 Korean film	Korean, Korean film, Han Fan
F26 Japanese film	Japan, Japanese film
F27 anime	cartoon, comic, anime, Conan

attribute feature levels. Table 2 finally summarizes the 27 categories of features mentioned in the user online reviews.

Formula 2 is used to calculate the user value attention to each feature film, statistics of users from the set of all reviews various types of key word frequency as molecules, respectively of the total number of key words to users as the denominator, calculate the proportion of the number of key, namely the

attention of users for each commodity characteristics, formation of user preference vectors.

$$\lambda_i = \frac{\sum U_{F_i}}{F_i} \quad (2)$$

3.4 Sentiment word score calculation

Sentimental tendency can be divided into three categories: positive, negative and neutral. The function of sentiment analysis is to analyze and infer the sentimental tendency of users. There are two main categories of mainstream sentimental tendency polarity discriminant methods: one is based on the dictionary discriminant method, the other is based on the review corpus discriminant method.

The dictionary discriminant method judges the polarity of the pending word by calculating the similarity between the pending word and the reference word in the standard semantic dictionary. In Chinese sentiment analysis, the most common way is to expand the sentiment dictionary based on the HowNet semantic dictionary or synonym word forest. The review corpus discriminant method relies on lexical cooccurrence frequency or modifier relationship for polarity judgment, such as PMI value and syntactic dependency. Currently, in the study of the identification of sentimental orientation, the artificial sentiment dictionary has a good application and can accurately distinguish the polarity of sentences and words, but it needs a lot of time and manpower to construct.

In the existing research, some scholars identify the sentiment polarity of the unincluded words with the help of the synonym word forest and explore the dependency relationship between the sentiment words and the evaluation objects through syntactic analysis. In order to achieve fine-grained sentiment analysis, some scholars have explored the influence of degree adverbs and their grades on sentimental intensity. This paper adopts the analysis method based on the HowNet sentiment dictionary to identify and calculate the polarity of sentiment words. The HowNet sentiment dictionary itself contains six subfiles, such as the "positive sentiment" thesaurus, the "negative sentiment" thesaurus, the "degree" level thesaurus, and the "proposition" thesaurus. However, online reviews are highly personalized, and many sentiment words are not included in the standard sentiment dictionary and cannot be recognized.

In the process of distinguishing the polarity of sentiment, the degree of closeness between the sentiment word and the reference word of positive meaning (or negative meaning) in the dictionary is calculated first and then the polarity of the sentiment word is determined by the semantic similarity. For a word "w", the polarity discrimination based on the HowNet emotion dictionary is shown in formula 3.

$$So - Hownet(w) = \sum_{i=1}^n Sim(w, com_i) - \sum_{i=1}^n Sim(w, der_i) \quad (3)$$

$Sim(w, com_i)$ represents the similarity between w and HowNet's commendatory term com_i , while $Sim(w, der_i)$ represents the similarity between w and HowNet's derogatory term der_i . If the calculated $so-Hownet(w)$ value is positive, then the polarity of the emotional word is positive. If it is negative, the polarity of emotion words is negative. If it is 0, then the emotion word polarity is neutral.

In order to realize fine-grained sentiment analysis, it is necessary not only to accurately judge the polarity of sentiment, but also to fully consider the modification function of degree adverbs on sentiment words to provide support for the classification of degree adverbs into analysis of different granularity. In order to distinguish the differences in the emotional intensity of users and take into account the influence of degree adverbs on emotional intensity and the effect of negative words on degree adverbs, this paper divides degree adverbs and negative adverbs into five levels, and sets the corresponding polarity according to the intensity of their emotional expression (see Table 3).

Based on the polarity coefficient of degree adverbs given in table 3, the words that do not contain degree adverbs are defined as basic sentiment words, and the words that contain degree adverbs are defined as compound sentiment words. Therefore, the words that conform to the compound sentiment words have the following three categories:

(1) Positive adverbs/negative adverbs + basic sentiment words, indicating the affirmation or negation of the part of speech of the feature words. The sentimental value calculation formula is shown in formula 4.

$$S_w = V_{w_i} * So - Hownet(w_i) \quad (4)$$

Table 3: Degree adverb polarity coefficient

Level	Category	Intensity	Example	Polarity coefficient
1	Positive adverb	Most	extremely	1.6
2	Positive adverb	Quite	especially, particularly	1.4
3	Positive adverb	Rather	further, rel- atively	0.8
4	Positive adverb	Slightly	a bit, a lit- tle	0.6
5	Negative adverb	Negative adverb	no,not	-1

S_w represents sentiment word score, $so - HowNet(wi)$ represents the score of basic sentiment word w_i , and V_{wi} represents degree adverb polarity coefficient.

(2) Positive adverb + negative adverb + basic sentiment word. After degree adverb, this form of negative word is the determination of negative action. The calculation formula for the score of sentiment shown in formula 5.

$$S_w = -(So - Hownet(wi)) * D_{wi} \quad (5)$$

S_w represents sentiment word score, $So - HowNet(wi)$ represents basic sentiment word score, and D_{wi} represents positive adverb polarity coefficient.

(3) Negative adverbs + positive adverbs + basic emotion words. A negative word before the positive adverb will weaken the negative degree of basic sentiment words. The calculation formula of sentiment words is shown in formula 6.

$$S_w = \eta * (So - Hownet(wi)) * D_{wi} \quad (6)$$

S_w represents the score of emotion words, η represents the degree of weakening of negative words, and its value is between 0 and 1. $So - HowNet(wi)$ represents the score of basic sentiment words w_i . D_{wi} represents the polarity coefficient of positive adverbs.

4 User recommendation based on feature emotion score

4.1 Calculation of sentimental value of feature words

Use the HowNet evaluation dictionary based on the sentimental polarity calculation method to determine the polarity of feature words, and use formulas 3, 4, 5 to calculate the sentimental value of compound sentimental words. Thus, the emotional score G_{kn} of user U on the Nth characteristic word of the Kth attribute feature of a specified item T is obtained. Considering the actual situation, users may use different words to express opinions on the same item features for many times in a review, so formula 7 is adopted to calculate the average score of user U on the f_k attribute features of the item.

$$S_k = \frac{\sum_{i=1}^N G_{kn}}{N} \quad (7)$$

According to formula 6, the sentimental score of item features in each user reviews can be obtained. By analyzing all the historical reviews of a certain item, the score of different attribute features of the item can be calculated. The score of the feature is expressed by the average score of sentiment of the feature. The calculation formula of the feature score is shown in formula 8. N represents the N reviews on an item, and R_k represents the average sentimental score on the Kth trait of an item.

$$R_k = \frac{\sum_{i=1}^N S_{ik}}{N} \quad (8)$$

4.2 User similarity calculation

In this paper, the cosine similarity method is adopted to calculate the similarity of user a and user b in the review text. $Sim(a, b)$ represents the similarity between user a and user b, R_a represents the review set of user a, R_b represents the review set of user b, K_a represents the feature set in user a's reviews, and $K_a \cap K_b$ represents the intersection of feature levels in user a and b's reviews.

$$sim(a, b) = \frac{\sum(R_a \cap R_b) * \sum(K_a \cap K_b)}{\sqrt{\sum R_a^2 * \sum K_a^2} * \sqrt{\sum R_b^2 * \sum K_b^2}} \quad (9)$$

4.3 User clustering

According to formula 7, the item sentiment score of users can be obtained. The Kmeans algorithm is used to cluster the item sentiment score and divide users into 7 categories. The user similarity calculated by formula 9 can be used to find the nearest neighbor of the target user's group.

User reviews with ID a5***7b are numerous and always provide high quality reviews. Therefore, "a5***7b" is selected as the target user to find its "nearest neighbor" according to the user similarity, and its nearest neighbor and similarity are shown in table 4.

Table 4: Recommended results

Serial number	UserID	similarity
1	73***41	0.6971
2	bc***11	0.4127
3	42***65	0.3795
...

4.4 Item recommendation generation

First, according to formula 7, the user's score on item features was calculated to obtain the user-feature score matrix, and then the similarity among users was obtained from formula 9 to obtain the K "nearest neighbors" of the target user. Then, the candidate items to be recommended are obtained according to the historical review of the "nearest neighbor", and the score of each feature of all candidate items is calculated by using formula 8. Finally, the personal preference of target users was integrated into equation 10 to predict the score of candidate items.

$$R = \sum_{k=1}^{27} R_k * \lambda_k \quad (10)$$

R_k is the predicted score of candidate projects, which is given by formula 8. λ_k is the target user's attention to the kth feature, given by formula 2. The final predicted score of each candidate project is sorted in descending order to form the final recommendation list and push to the target user, as shown in table 5.

5 Experiment results and analysis

5.1 Data preparation

In order to verify the actual effect of personalized recommendation based on emotion analysis, this paper adopts Scrapy method based on python to obtain movie information of douban, including user ID, douban ID, user review and user score. The collected data set includes 215032 reviews, delete users with less than 10 movies and all movies with less than 10 reviews in the data table, and get a total of 2283 users' 67376 reviews on 1464 movies.

Table 5: Recommended results

Serial number	MovieID	Predicted score
1	10514820	1.695
2	4824996	1.639
3	26592180	1.596
4	6833818	1.479
5	1308575	1.479
6	1444533	1.421
7	3718526	1.369
...

5.2 Assessment indicator

Predictive accuracy P represents the probability that the user may like an item in the recommendation list, which can show the accuracy of the recommender system. The formula to calculate predictive accuracy of recommender system is as follow:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (11)$$

Recall rate R represents the proportion of items users like in the recommendation list, which can show users' satisfaction degree with the recommendation results. The higher the recall rate is, the higher satisfaction degree users have. The formula to calculate the recall rate of recommender system is as follow:

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (12)$$

5.3 Result analysis

(1)Extraction effect of feature - sentiment words pairs

Taking Douban film review data as an example, based on previous studies, this paper proposes to employ user review data for sentiment analysis to explore features of films and preferences of individual users, improve feature and emotion word extraction methods, and summarize four syntactic dependencies. Based on the four syntactic dependencies, the user reviews are analyzed with fine granularity, which can effectively improve the extraction accuracy of feature and sentiment words.

In order to verify the effectiveness of the feature-sentiment words pairs extraction method proposed in this paper, 1000 reviews were randomly selected from the review data set, and feature words and sentiment words were manually labeled, resulting in 75 feature words and 113 sentiment words. Then, using the method proposed by Hu and Liu as the benchmark, the extraction method of feature - sentiment words pairs proposed in this paper is compared.

In this paper, accuracy and recall rate are selected as evaluation indexes, in which accuracy refers to the percentage of extracted correct feature words and sentiment words, and recall rate refers to the percentage of extracted correct feature words and emotion words in labeled words. The experimental results are shown in tables 6 and 7.

Table 6: Feature word extraction results

methods	accuracy rate /%	recall rate /%
Hu	52.11	37.25
Extraction method in this paper	53.76	63.44

(2)Recommendation effect analysis By analyzing users' online reviews, this paper classifies movie features into 27 categories to further refine movie features and then statistically obtains users' attention to various attribute features from users' own review collection to form user preferences. User preference

Table 7: Emotional word extraction results

methods	accuracy rate /%	recall rate /%
Hu	52.11	37.25
Extraction method in this paper	53.76	63.44

is integrated to predict the score of recommended items, find the most consistent with user preference and generate a recommendation list.

The traditional recommendation algorithm based on score only identifies as the nearest neighbor the users with the same score as the target users and then recommends the favorite items to the target users while ignoring the real preference of the users. Recommendation accuracy is thus low. However, the recommendation model proposed in this paper considers the implicit preference mined from user reviews and shows a good recommendation effect. Every recommended item can meet the preference requirements of users in some aspect. Compared with the traditional recommendation algorithm based on score, the recommendation method in this paper is obviously improved in accuracy and recall rate. The results of the precision and recall rates of user-based collaborative filtering, project-based collaborative filtering and the proposed methods in this paper are shown in table 8.

Table 8: Feature word extraction results

methods	accuracy rate /%	recall rate /%
UserCF	22.45	22.50
ItemCF	21.44	22.39
Recommended algorithm in this paper	25.53	26.49

(3) Degree adverb polarity intensity quantification

This study shows that the polarity of emotion words is easy to determine and that degree adverbs represent the degree of words or sentence which can strengthen or weaken the strength of emotion words in the application of emotion analysis. In addition, if the order of degree adverbs and negative words is different, the polarity of emotion words expressed will be greatly different. This paper gives full consideration to the influence of degree adverbs on emotional intensity, defines five polarity coefficients of degree adverbs, and considers the co-occurrence order of degree adverbs and negative words. In chapter 3.4, a somewhat innovative set of computing standards for emotional intensity that conforms to the general expression of users is designed to conduct fine-grained emotional scores and improve the accuracy of emotional analysis of user reviews.

(4) User clustering effect

The recommendation model in this paper belongs to the user-based collaborative filtering approach but is different from traditional user-based collaborative filtering in that it fully considers the user preference information implied in user reviews. Traditional methods cluster users based on ratings only so that two users with completely different preferences who give the same ratings are clustered together. For example, users who habitually give favorable reviews are always grouped together even though their actual preferences and needs are very different. In this paper, a novel recommendation model based on emotion analysis is proposed. Based on the emotional score of movie features, users are clustered, and users with the same emotional preference are gathered to form a user group. Through the emotional analysis of user reviews, the fine-grained preferences of users for specific commodities are obtained for clustering and the similarity of the overall score of items by users in the traditional collaborative filtering is transformed into similarity based on the attributes and characteristics of items. At the same time, the accuracy of recommendation is improved by considering the preferences of users. By selecting the historical videos of other users with high similarity in the user group, calculating the scores of all candidate items with various features, and integrating the personal preferences of target users, group recommendation in the process of personalized recommendation is achieved. The user clustering effect has a direct impact on the movie recommendation. Therefore, the method in this paper indirectly improves the recommendation effect by effectively improving the user clustering effect.

6 Conclusion

Due to the ever-increasing number of people online and the diversity of Internet activity, research on improving the accuracy of personalized recommendation can be endless. However, the current recommendation method only explores the individual needs of a user based on their past purchase behaviors and ratings and ignores personal preferences expressed in online reviews. User reviews have become a growing information carrier. It is a new perspective for personalized recommendation research to deeply explore user review content with emotion analysis as the main means.

This paper proposes a personalized recommendation model based on emotion analysis to solve the problem of information overload of users' online reviews. Taking Douban movies as an example, this paper proposes the analysis of user reviews to explore features of movies and preferences of individual users. In addition, this paper improves the extraction methods of feature words and emotion words and summarizes four syntactic dependencies. Based on the four syntactic dependencies, it carries out fine-grained analysis of user reviews to effectively improve the extraction accuracy of feature words and emotion words. In addition, considering the co-occurrence order of degree adverbs and negative words, a set of computing standards for emotional intensity that conforms to the conventional expression of users is designed to calculate fine grained emotional scores and improve the accuracy of emotional analysis of user reviews. A novel recommendation model based on emotion analysis is proposed. The model converts the similarity of the overall score of items in traditional collaborative filtering to the similarity of the score based on the attributes of items, improves the clustering effect of users, and considers the preferences of users to improve the accuracy of recommendation. Based on the emotional score of the movie features, the users were clustered, and the users with the same emotional preference were gathered to form a user group. Group recommendation was realized in the personalized recommendation process to improve the recommendation efficiency.

Funding

This work was funded by Research on scene behavior intelligence in collaborative network based on data analysis (NSFC 61972029).

Author contributions. Conflict of interest

The authors contributed equally to this work. The authors declare no conflict of interest.

References

- [1] Amplayo, R.K.; Song, M..(2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews, *Data and Knowledge Engineering*. 110, 54-67, 2017
- [2] Chakraborty, K.; Bhattacharyya, S.; Bag, R. ; Hassanien, A.A.(2019). Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Techniques, *Social Network Analytics*, 127-147, 2019.
- [3] Chang, D.; Gui, H. Y.; Fan, R.(2019). Application of Improved Collaborative Filtering in the Recommendation of E-commerce Commodities, *International Journal of Computers Communications and Control*, 14(4), 489-502, 2019.
- [4] Chen, J.; Kou, G.; Peng, Y.(2018). The dynamic effects of online product reviews on purchase decisions, *Technological and Economic Development of Economy*, 24(5), 2045-2064, 2018.
- [5] Deng, Z.; Luo, K.; Yu, H.(2014). A study of supervised term weighting scheme for sentiment analysis, *WExpert Systems with Application*, 41(7), 3506-3513, 2014.
- [6] Du, C.; Huang, L. (2019). Sentiment Analysis Method Based On Piecewise Convolutional Neural Network and Generative Adversarial Network, *International Journal of Computers Communications & Control*, 14(1),7-20, 2019.

- [7] Gan, Q.; Ferns, B.H.; Yu, Y.; Jin, L.(2017). A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews, *Journal of Quality Assurance in Hospitality and Tourism*, 18(4), 465-492, 2017.
- [8] Gifu, D; Cioca, M.(2019). Detecting Emotions in Comments on Forums, *International Journal of Computers Communications & Control*, 9(6), 694-702, 2019.
- [9] Ghose, A.; Ipeirotis, P.G.(2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews, *Proceedings of the Ninth International Conference on Electronic Commerce. ACM*, 303-310, 2007.
- [10] Goldberg, D.; Nichols, D.; Oki, B.M. et al. (1992). Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM*, 35(12), 61-70, 1992.
- [11] Hu, M.Q.; Liu, B.(2004). Mining opinion features in customer reviews, *In Proc. of the 19th National Conf,on Artificial Intelligence(AAAI 2004),AAAI Press*, 755-760, 2004.
- [12] Hu, M.; Liu, B.(2004). Mining and summarizing customer reviews, *Proceedings of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.New York:ACM Press*, 168-177, 2004.
- [13] Liu, H.; He, J.; Wang, T.; Song, W.; Du, X. (2013). Combining user preferences and user opinions for accurate recommendation, *Electronic Commerce Research and Applications*, 12(1), 14-23, 2013.
- [14] Manek, A. S.; Shenoy, P. D.; Mohan, M. C.(2017). Aspect Term Extraction for Sentiment Analysis in Large Movie Reviews Using Gini Index Feature Selection Method and SVM Classifier, *World Wide Web-internet 2004:Web Information Systems*, 20(2), 135-154, 2017.
- [15] Mudambi, S. M.; Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon. com , *MIS Quarterly*, 34 (1): 185-200,2010.
- [16] Perikos, I.;Hatzilygeroudis, I.(2016). Recognizing emotions in text using ensemble of classifiers, *Engineering Applications of Artificial Intelligence*, 51, 191-201, 2016.
- [17] Rao, Y.H.; Lei, J.S.; Liu, W.Y.; Li, Q.; Chen, M.L.(2014). Building emotional dictionary for sentiment analysis of online news, *World Wide Web: Internet and Web Information Systems*, 17(4), 723-742, 2014.
- [18] Sayeedunnessa, S F.; Hussain, A. R.; Hameed, M. A.(2013). Supervised opinion mining of social network data using a bag-of-words approach on the cloud, *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Application(BIC-TA2012). India:Springer*, 299-309, 2013.
- [19] Venkata, G.; Sreevatsava, R.(2019). An Effective Study on Sentiment Analysis Using Machine Learning Approach, *International Journal of Management, Technology And Engineering*, 19(1), 9, 2009.
- [20] Wang, Y.; Liu, Y.; Yu, X.(2012). Collaborative Filtering with Aspect-Based Opinion Mining: A Tensor Factorization Approach, *Proceedings of the IEEE International Conference on Data Mining, Piscataway NJ:IEEE*, 1152-1157, 2012.
- [21] Wang, W.; Wang, H.(2015). Opinion-enhanced Collaborative Filtering for Recommender Systems Sentiment Analysis, *New Review in Hypermedia and Multimedia*, 21(3-4), 1-23, 2015.
- [22] [Online]. Available: www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwtjbg/201908/.html



Copyright ©2020 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Chen, S.; Lv, X.; Gou J.(2020). Personalized Recommendation Model: An Online Comment Sentiment Based Analysis, *International Journal of Computers Communications & Control*, 15(1), 1004, 2020.
<https://doi.org/10.15837/ijccc.2020.1.3764>